MBMG
Metabarcoding & Metagenomics

Metabarcoding and Metagenomics 6: 239–244
DOI 10.3897/mbmg.6.84960

Forum Paper

Introducing guidelines for publishing DNA-derived occurrence data through biodiversity data platforms

R. Henrik Nilsson[1], Anders F. Andersson[2], Andrew Bissett[3], Anders G. Finstad[4], Frode Fossøy[5], Marie Grosjean[6], Michael Hope[7], Thomas S. Jeppesen[6], Urmas Kõljalg[8], Daniel Lundin[9], Maria Prager[10,11], Saara Suominen[12], Cecilie S. Svenningsen[13], Dmitry Schigel[6]

1 *University of Gothenburg, Department of Biological and Environmental Sciences, Gothenburg Global Biodiversity Centre, Box 461, 405 30 Göteborg, Sweden*
2 *Science for Life Laboratory, Department of Gene Technology, KTH Royal Institute of Technology, 17121 Stockholm, Sweden*
3 *CSIRO O&A, GPO box 1533, Hobart, Tasmania, 7000, Australia*
4 *Department of Natural History, Centre for Biodiversity Dynamics, Norwegian University of Science and Technology, Trondheim, Norway*
5 *Norwegian institute for nature research (NINA), P.O. Box 5685 Torgarden, NO-7485 Trondheim, Norway*
6 *Global Biodiversity Information Facility (GBIF), Secretariat, Universitetsparken 15, 2100 København Ø, Denmark*
7 *Atlas of Living Australia, CSIRO National Collections & Marine Infrastructure, GPO Box 1700, Canberra ACT 2601, Australia*
8 *Natural History Museum and Botanical Garden, University of Tartu, 46 Vanemuise Street, 51003 Tartu, Estonia*
9 *Centre for Ecology and Evolution in Microbial model Systems - EEMiS, Linnaeus University, SE-39182 Kalmar, Sweden*
10 *Science for Life Laboratory, Department of Ecology, Environment and Plant Sciences, Stockholm University, Stockholm, Sweden*
11 *Department of Microbiology, Tumor and Cell Biology, Karolinska Institutet, Solna, Sweden*
12 *UNESCO Intergovernmental Oceanographic Commission (IOC), Ocean Biodiversity Information System (OBIS), IOC Project Office for IODE, Oostende, Belgium*
13 *Natural History Museum of Denmark, University of Copenhagen, Øster Voldgade 5-7, 1350 Copenhagen, Denmark*

Corresponding author: Dmitry Schigel (dschigel@gbif.org)

Academic editor: Dirk Steinke | **Received** 6 April 2022 | **Accepted** 5 July 2022 | **Published** 2 August 2022

Abstract

DNA sequencing efforts of environmental and other biological samples disclose unprecedented and largely untapped opportunities for advances in the taxonomy, ecology, and geographical distributions of our living world. To realise this potential, DNA-derived occurrence data (notably sequences with dates and coordinates) – much like traditional specimens and observations – need to be discoverable and interpretable through biodiversity data platforms. The Global Biodiversity Information Facility (GBIF) recently headed a community effort to assemble a set of guidelines for publishing DNA-derived data. These guidelines target the principles and approaches of exposing DNA-derived occurrence data in the context of broader biodiversity data. They cover a choice of terms using a controlled vocabulary, common pitfalls, and good practices, without going into platform-specific details. Our hope is that they will benefit anyone interested in better exposure of DNA-derived occurrence data through general biodiversity data platforms, including national biodiversity portals. This paper provides a brief rationale and an overview of the guidelines, an up-to-date version of which is maintained at https://doi.org/10.35035/doc-vf1a-nr22. User feedback and interaction are encouraged as new techniques and best practices emerge.

Key Words

biological data management, DNA sequences, metabarcoding, metagenomics, occurrence record, open data, scientific credit, scientific reproducibility

Introduction

The last 30 years have brought an increased understanding of the immense power of molecular methods for documenting the diversity of life on earth. DNA-derived data enable us to also record inconspicuous and even undescribed species – taxa that typically fall below the radar of vetted protocols for field work, checklists, and depositions into natural science collections. Expanding the concept of biological occurrences to routinely include molecular detections is a hotly discussed topic that has only relatively recently moved beyond the conceptual stage, through the Global Biodiversity Information Facility's (GBIF; www.gbif.org/) inclusion of fungal molecular occurrence data (Nilsson et al. 2019). Other initiatives soon followed, and yet others are in the planning stage. This puts us at a pivotal time in the history of biology; by reaching agreement on how we should record and report on an organism as present in some substrate or locality through molecular data, we can hopefully avoid issues of data heterogeneity and incomparability that have plagued other scientific fields for decades (Leebens-Mack et al. 2006; Yilmaz et al. 2011). Moreover, clear documentation of the computational processing from raw DNA sequence data to deduced species observations will improve interoperability and scientific reproducibility, including subsequent data reanalysis using improved methods, as put forth through the FAIR principles (Wilkinson et al. 2016). In 2020, the present authors – an international community headed by GBIF – set out to produce a guidelines document for standardised and reproducible generation and representation of biological occurrences through molecular data. August 2021 saw the first release of this guide (https://doi.org/10.35035/doc-vf1a-nr22), whose foundational principles are outlined below.

Data packing and mapping

Our goal was to make the guide comprehensive enough to cover at least the most popular of the many DNA-based approaches used to characterise the world's biota, with a primary focus on metabarcoding, metagenomics, and quantitative PCR (qPCR and ddPCR). The guide assumes the data to have been collected, processed, and analysed in appropriate ways (Bustin et al. 2009; Budowle et al. 2014; Thalinger et al. 2021; Tedersoo et al. 2022). We sought to put forth a set of instructions on how to format data from DNA metabarcoding, metagenomics, and qPCR/ddPCR projects into datasets amenable to algorithmic interpretation by the major biodiversity informatics platforms. In this process we wanted to reflect the Darwin Core standard (DwC; Wieczorek et al. 2012), a controlled vocabulary intended to facilitate the sharing of information about biological diversity by providing identifiers, labels, and definitions. This essentially meant that we needed to specify which 'core' and 'extension' files to use, and how to best map specific data items to extant (or novel) DwC terms. Recognising that at least some metadata, such as processed barcode sequences, would need to point to individual occurrences, we settled for an Occurrence core. Choosing an extension file to hold sequence-related metadata was less straightforward, as several related initiatives already existed within the DwC community, although their output did not fully cater to all our needs. These available initiatives included the GGBN Amplification and the MIxS Sample specification, provided by the Global Genome Biodiversity Network and the Genomic Standards Consortium (GSC), respectively. As the latter format derived from the same GSC family of minimum information standards used in sample registration and raw sequence archiving with the European Nucleotide Archive (ENA; Cummins et al. 2022), it was selected as a starting point for our extension file.

The mapping process started with a spreadsheet comparison of (meta)data fields used in a selection of sequence-based datasets provided by GBIF including, e.g., output from the MGnify pipeline (Mitchell et al. 2020) and the Biowide project (Frøslev and Ejrnæs 2018), as well as by the present authors (e.g., Finstad et al. 2020). These fields were mapped to fields both in the MIxS sample specification and in the more extensive body of GSC MIxS checklists. The objectives were to: 1) identify ambiguities, i.e., cases where the same information type was given in different fields; 2) discuss where we could make use of standard DwC or MIxS fields, and where we needed to add novel fields, if any; and 3) define sets of recommended and required fields for this specific type of occurrence data.

Blending individual elements from existing standards may risk jeopardising universality and inclusiveness of detail in the resulting mix but should improve interoperability and maximise the coverage of cases across biomes (the minimum standard approach, see Rund et al. 2019). We participated in a joint Biodiversity Information Standards and Genomic Standards Consortium endeavour to align efforts on the DwC and MIxS specifications. This work included semantic mapping between DwC/MIxS terminologies, harmonised use of identifiers, and test ingestions of metabarcoding datasets using the proposed DwC extension. These results are presented in Meyer et al. (2021) and benefited the present guidelines.

At the time of writing, none of GBIF, OBIS, or ALA is capable of directly ingesting biological samples from observation (taxon/operational taxonomic unit) contingency tables. Therefore, the mapping step in Fig. 1 also implies conversion into Darwin Core archives. We are not aware of any standard tools for this conversion, so some degree of scripting is involved. The conversion can be described as follows: for each taxon/operational taxonomic unit (OTU; Blaxter et al. 2005) in a sample, write one row to the Occurrence and DNA derived data CSV files that combines information from the Sample and Taxon/OTU metadata, including dates and coordinates. CSV headers in the Occurrence and DNA derived data CSV files should follow the recommendations in the guide (https://

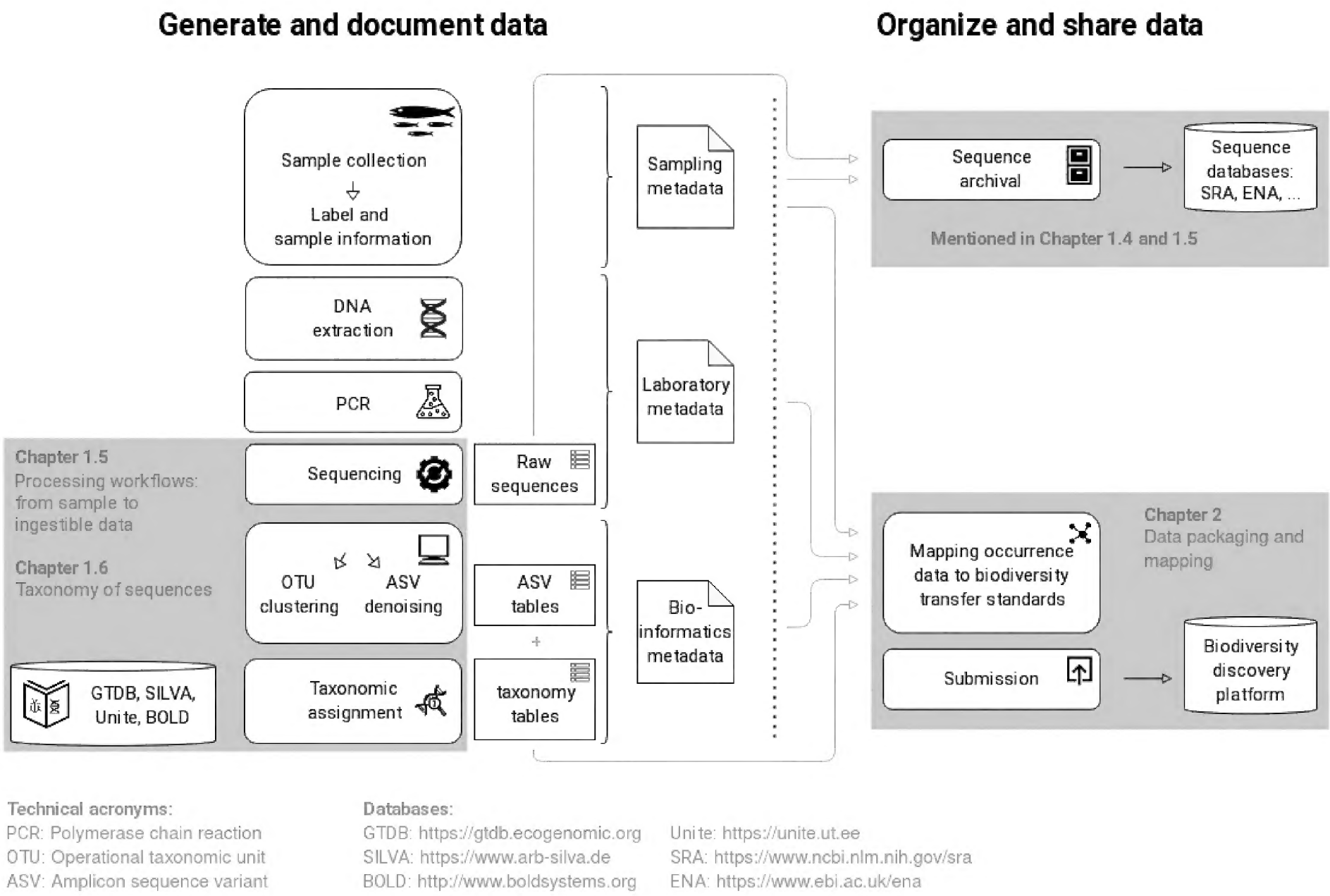

Figure 1. Overall workflow for DNA sequence-derived biodiversity data as described in the guide (https://doi.org/10.35035/doc-vf1a-nr22). Chapter numbers refer to chapters in the guide.

doi.org/10.35035/doc-vf1a-nr22) sections 2.2.1 (metabarcoding (eDNA) and barcoding data) and 2.2.2 (ddPCR / qPCR data). The CSV files should be packaged into a Darwin Core archive along with an Ecological Metadata Language file (EML; Jones et al. 2019) describing the dataset, as shown in Fig. 2. Preparing the EML and mapping CSV column headers can be done in IPT (Robertson et al. 2014), which supports registering datasets for ingestion into GBIF. General information on the anatomy of Darwin Core archives can be found at https://dwc.tdwg.org/text/.

Outline of the guidelines

The guide is maintained at https://doi.org/10.35035/doc-vf1a-nr22. We intend it to be a living document that is updated as new techniques and best practices emerge, and for this reason the guide is not presented in a static version in the present publication. An overview of the aspects covered by the guide is provided in Fig. 1. The guide starts at the raw data step and covers various aspects of data treatment and analysis. It touches upon deposition of raw data into appropriate repositories and then pays particular attention to turning the raw sequence data into Darwin Core Archive (DwC-A; https://dwc.tdwg.org/) compatible, enriched occurrence record files for data publication in, e.g., the GBIF and Atlas of Living Australia (ALA; Belbin et al. 2021) networks and the Ocean Biodiversity information system (OBIS; https://obis.org). Other major networks and national biodiversity data platforms are currently exploring applicability of the guide to their needs. The guide wraps up by outlining practical aspects of data publication and dissemination (Fig. 2). The three chapters of the guide comprise numerous best practises and a total of 99 proposed (meta)data fields for the various data types covered by the guide. The relevance of the various components of the guide will vary with the specifics of the study and data at hand, but our intention is that the guide should provide both a framework and flexibility to additionally accommodate study types unforeseen by us.

Discussion and outlook

The nature of the stakeholders of biodiversity data platforms is very diverse. Users and data depositors include students, researchers, biodiversity data managers, governmental and private agencies, policy makers, and bioinformaticians. Not all stakeholders are perhaps in the habit of approaching biological evidence through molecular means, but we sense that the interest in exploring DNA-derived data through biodiversity data platforms is growing steadily. While this highlights the need for a set of guidelines and recommendations of the present kind, it also suggests that situations and cases unforeseen by the authors and contributors of this guide are likely to surface. Similarly,

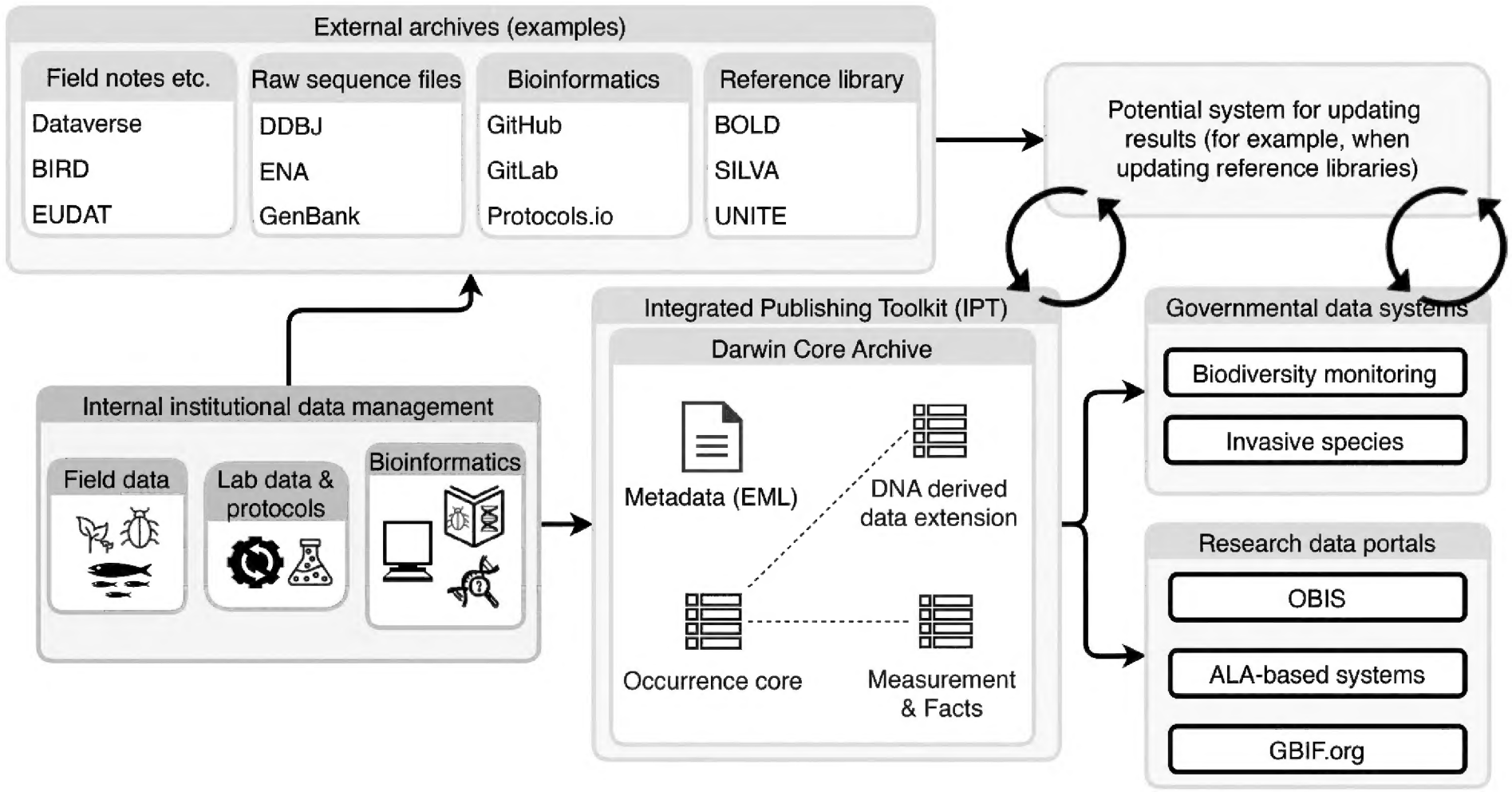

Figure 2. Outline of a platform for reporting and publishing DNA sequences and associated metadata (green box) based on existing systems and data standards (grey boxes). An envisioned system for regular (based on machine-to-machine reading of data) update of results (white box) can either read, and update, the Darwin Core Archive or various other administration systems. The data transfer between the various elements (black arrows) will require various degrees of data transformation and harmonisation and may include either mechanical or human quality assessment. The items "DNA-derived data extension" and "Measurements & Facts" refer to data that must, should, or could be bundled with occurrence data and are detailed in section 2.2 of the guide.

recommendations and best practices are likely to change over time as new techniques and approaches emerge. We are, for instance, in the process of considering resources such as the BOLD Handbook, the Biological Observation Matrix (BIOM) format, and the EDAM ontology of bioscientific data analysis and data management (http://edamontology.org/page). Similarly, data formats that support more complex relational and hierarchical data – notably the Frictionless Data Format – are interesting and very relevant developments for the study of biodiversity. The guide has already seen a number of minor updates and improvements since its formal August 2021 release, and our ambition is to keep it updated over time. User feedback is a crucial component of this endeavour, and we warmly welcome user interaction at the URL provided in the Results section.

The purpose of exposing DNA-derived occurrence data through biodiversity platforms is to enable reuse of these data alongside other biodiversity data types. Connecting DNA sequences to traditional nomenclature through voucher specimen sequencing is still in progress in genetic reference databases. Indeed, recording sequences alongside occurrences will allow continuous update and reconfirmation of taxonomic classifications. To facilitate comparisons to traditional observations, links to databases of scientific names should be maintained. For example, OBIS adopted the present guidelines and additionally requires a direct link with Linnean names through the World Register of Marine Species (WoRMS; https://www.marinespecies.org) catalogue. Indeed, through the development and adoption of these guidelines through multiple biodiversity data networks, the sharing of large amounts of data arising from genetic studies will be made easier and promote wider use of those data. Future plans include work to enable publishing datasets across both GBIF and OBIS through a single data submission instance.

A hurdle towards the goal of integrating DNA-based occurrences into routine biological practice is the somewhat poor track record of biology when it comes to making actual research data available to begin with (e.g., Hinchliff et al. 2015; Khan et al. 2021). The final, published research paper is all too often seen as the sole end product of the research project in question, and little, if any, effort is made to facilitate re-interpretation and re-use of the underlying results and data for the same or other purposes (Durkin et al. 2020; Abarenkov et al. 2022). Our understanding of our living world comes out at the losing end of decisions of this kind, and we are happy to note an incipient trend towards increasing awareness of the role of data in biology (Penev et al. 2017; Mandeville et al. 2021). There are many reasons why it makes scientific and professional sense to report DNA-derived occurrence data in an open and reproducible way. Notably, it contributes to taxonomic and ecological advances, it highlights taxa concerned in the context of biological conservation, it may invite unexpected collaborations, it is increasingly being favoured by research councils and other funding bodies, and it is likely to increase citations (Culina et al. 2018; Christensen et al. 2019; Khan et al. 2021). Additionally, it also provides a mechanism to store occurrence records of undescribed species (Kõljalg et al. 2020). When these yet-to-be-described taxa are finally linked to new Linnaean names, all these linked occurrence records will be

immediately available (Nilsson et al. 2019). Each of these benefits provides a strong rationale for professionals to adopt the practices outlined in this guide, helping them to highlight a significant proportion of biodiversity, speed up its discovery and formal description, and integrate it into biological conservation and policymaking (Scholz et al. 2022). Biology is perhaps a field that has been slow to fully adopt the concepts of data publishing and reuse, but we hope that the present guide will contribute to the popularisation and perhaps standardisation of new biological data types. We certainly anticipate a near future where biological occurrences are routinely pursued in light of both traditional and molecular data for all groups of organisms.

Funding

The participation of AFA, DL, and MP in this project was partly funded through the Swedish Biodiversity Data Infrastructure (SBDI) funded by its partner organisations and the Swedish Research Council VR through Grant No 2019-00242.

Competing interests

The authors have declared that no competing interests exist.

Acknowledgements

Valuable discussions with members of the ELIXIR, iBOL, GGBN, GLOMICON, and OBIS networks contributed to compilation of this draft. We are especially grateful for input and encouragement from Kessy Abarenkov, Andrew Bentley, Matt Blissett, Pier Luigi Buttigieg, Kyle Copas, Camila A. Plata Corredor, Gabriele Dröge, Torbjørn Ekrem, Tobias Guldberg Frøslev, Birgit Gemeinholzer, Quentin Groom, Tim Hirsch, Donald Hobern, Hamish Holewa, Corinne Martin, Raissa Meyer, Chris Mungall, Daniel Noesgaard, Corinna Paeper, Pieter Provoost, Tim Robertson, Maxime Sweetlove, Andrew Young, John Waller, Ramona Walls, John Wieczorek, and Lucie Zinger who contributed to the GBIF community review process. We finally acknowledge the important role of Andrew Young in instigating the guidelines effort. An anonymous reviewer is acknowledged for providing valuable feedback on an earlier draft of the manuscript.

References

Abarenkov K, Kristiansson E, Ryberg M, Nogal-Prata S, Gómez-Martínez D, Stüer-Patowsky K, Jansson T, Põlme S, Ghobad-Nejhad M, Corcoll N, Scharn R, Sánchez-García M, Khomich M, Wurzbacher C, Nilsson RH (2022) The curse of the uncultured fungus. MycoKeys 86: 177–194. https://doi.org/10.3897/mycokeys.86.76053

Belbin L, Wallis E, Hobern D, Zerger A (2021) The Atlas of Living Australia: History, current state and future directions. Biodiversity Data Journal 9: e65023. https://doi.org/10.3897/BDJ.9.e65023

Blaxter M, Mann J, Chapman T, Thomas F, Whitton C, Floyd R, Abebe E (2005) Defining operational taxonomic units using DNA barcode data. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences 360(1462): 1935–1943. https://doi.org/10.1098/rstb.2005.1725

Budowle B, Connell ND, Bielecka-Oder A, Colwell RR, Corbett CR, Fletcher J, Forsman M, Kadavy DR, Markotic A, Morse SA, Murch RS, Sajantila A, Schmedes SE, Ternus KL, Turner SD, Minot S (2014) Validation of high throughput sequencing and microbial forensics applications. Investigative Genetics 5(1): 1–18. https://doi.org/10.1186/2041-2223-5-9

Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, Mueller R, Nolan T, Pfaffl MW, Shipley GL, Vandesompele J, Wittwer CT (2009) The MIQE guidelines: Minimum information for publication of quantitative real-time PCR experiments. Clinical Chemistry 55(4): 611–622. https://doi.org/10.1373/clinchem.2008.112797

Christensen G, Dafoe A, Miguel E, Moore DA, Rose AK (2019) A study of the impact of data sharing on article citations using journal policies as a natural experiment. PLoS ONE 14(12): e0225883. https://doi.org/10.1371/journal.pone.0225883

Culina A, Baglioni M, Crowther TW, Visser ME, Woutersen-Windhouwer S, Manghi P (2018) Navigating the unfolding open data landscape in ecology and evolution. Nature Ecology & Evolution 2(3): 420–426. https://doi.org/10.1038/s41559-017-0458-2

Cummins C, Ahamed A, Aslam R, Burgin J, Devraj R, Edbali O, Gupta D, Harrison PW, Haseeb M, Holt S, Ibrahim T, Ivanov E, Jayathilaka S, Kadhirvelu V, Kay S, Kumar M, Lathi A, Leinonen R, Madeira F, Madhusoodanan N, Mansurova M, O'Cathail C, Pearce M, Pesant S, Rahman N, Rajan J, Rinck G, Selvakumar S, Sokolov A, Suman S, Thorne R, Totoo R, Vijayaraja S, Waheed Z, Zyoud A, Lopez R, Burdett T, Cochrane G (2022) The European Nucleotide Archive in 2021. Nucleic Acids Research 50(D1): D106–D110. https://doi.org/10.1093/nar/gkab1051

Durkin L, Jansson T, Sanchez M, Khomich M, Ryberg M, Kristiansson E, Nilsson RH (2020) When mycologists describe new species, not all relevant information is provided (clearly enough). MycoKeys 72: 109–128. https://doi.org/10.3897/mycokeys.72.56691

Finstad AG, de Boer H, Brurberg MB, Dahle G, Skarsfjord Edgar K, Eiler A, Ekrem T, Endresen D, Fossøy F, Hansen H, Hobæk A, Hoem SA, Hosia A, Hovstad KA, Stjernegaard Jeppesen T, Johnsen A, Kallioiniemi E, Larsen A, Lifjeld JT, Pitelkova I, Prager M, Louise Ray J, Salvesen I, Vrålstad T, Willassen R (2020) Kriterier for lagring av miljø-DNA prøver og data, herunder henvisning til referansemateriale. Miljødirektoratet rapport M-1638: 1–40. https://www.miljodirektoratet.no/globalassets/publikasjoner/m1638/m1638.pdf

Frøslev T, Ejrnæs R (2018) BIOWIDE eDNA Fungi dataset. Danish Biodiversity Information Facility. Occurrence dataset https://doi.org/10.15468/nesbvx [accessed via GBIF.org on 2022-03-14]

Hinchliff CE, Smith SA, Allman JF, Burleigh JG, Chaudhary R, Coghill LM, Crandall KA, Deng J, Drew BT, Gazis R, Gude K, Hibbett DS, Katz LA, Laughinghouse HD IV, McTavish EJ, Midford PE, Owen CL, Ree RH, Rees JA, Soltis DE, Williams T, Cranston KA (2015) Synthesis of phylogeny and taxonomy into a comprehensive tree of life. Proceedings of the National Academy of Sciences of the United States of America 112(41): 12764–12769. https://doi.org/10.1073/pnas.1423041112

Jones MB, O'Brien M, Mecum B, Boettiger C, Schildhauer M, Maier M, Whiteaker T, Earl S, Chong S (2019) Ecological Metadata Language version 2.2.0. KNB Data Repository.

Khan N, Thelwall M, Kousha K (2021) Measuring the impact of biodiversity datasets: Data reuse, citations and altmetrics. Scientometrics 126(4): 3621–3639. https://doi.org/10.1007/s11192-021-03890-6

Kõljalg U, Nilsson HR, Schigel D, Tedersoo L, Larsson K-H, May TW, Taylor AFS, Jeppesen TS, Frøslev TG, Lindahl BD, Põldmaa K, Saar I, Suija A, Savchenko A, Yatsiuk I, Adojaan K, Ivanov F, Piirmann T, Pöhönen R, Zirk A, Abarenkov K (2020) The taxon hypothesis paradigm– On the unambiguous detection and communication of taxa. Microorganisms 8(12): 1910. https://doi.org/10.3390/microorganisms8121910

Leebens-Mack J, Vision T, Brenner E, Bowers JE, Cannon S, Clement MJ, Cunningham CW, DePamphilis C, DeSalle R, Doyle JJ, Eisen JA, Gu X, Harshman J, Jansen RK, Kellogg EA, Koonin EV, Mishler BD, Philippe H, Pires JC, Qiu Y-L, Rhee SY, Sjölander K, Soltis DE, Soltis PS, Stevenson DW, Wall K, Warnow T, Zmasek C (2006) Taking the first steps towards a standard for reporting on phylogenies: Minimum Information About a Phylogenetic Analysis (MIAPA). OMICS: A Journal of Integrative Biology 10(2): 231–237. https://doi.org/10.1089/omi.2006.10.231

Mandeville CP, Koch W, Nilsen EB, Finstad AG (2021) Open data practices among users of primary biodiversity data. Bioscience 71(11): 1128–1147. https://doi.org/10.1093/biosci/biab072

Meyer R, Buttigieg PL, Wieczorek J, Stjernegaard Jeppesen T, Duncan WD, Gan Y-M, Sweetlove M, Suominen S (2021) Aligning standards communities: Sustainable Darwin Core MIxS interoperability. Biodiversity Information Science and Standards 5: e73775. https://doi.org/10.3897/biss.5.73775

Mitchell AL, Almeida A, Beracochea M, Boland M, Burgin J, Cochrane G, Crusoe MR, Kale V, Potter SC, Richardson LJ (2020) MGnify: The microbiome analysis resource in 2020. Nucleic Acids Research 48(D1): D570–D578. https://doi.org/10.1093/nar/gkz1035

Nilsson RH, Larsson KH, Taylor AFS, Bengtsson-Palme J, Jeppesen TS, Schigel D, Kennedy P, Picard K, Glöckner FO, Tedersoo L, Saar I, Kõljalg U, Abarenkov K (2019) The UNITE database for molecular identification of fungi: Handling dark taxa and parallel taxonomic classifications. Nucleic Acids Research 47(D1): D259–D264. https://doi.org/10.1093/nar/gky1022

Penev L, Mietchen D, Chavan VS, Hagedorn G, Smith VS, Shotton D, Tuama ÉÓ, Senderov V, Georgiev T, Stoev P, Groom QJ, Remsen D, Edmunds SC (2017) Strategies and guidelines for scholarly publishing of biodiversity data. Research Ideas and Outcomes 3: e12431. https://doi.org/10.3897/rio.3.e12431

Robertson T, Döring M, Guralnick R, Bloom D, Wieczorek J, Braak K, Otegui J, Russell L, Desmet P (2014) The GBIF Integrated Publishing Toolkit: Facilitating the efficient publishing of biodiversity data on the Internet. PLoS ONE 9(8): e102623. https://doi.org/10.1371/journal.pone.0102623

Rund SSC, Braak K, Cator L, Copas K, Emrich SJ, Giraldo-Calderón GI, Johansson MA, Heydari N, Hobern D, Kelly SA, Lawson D, Lord C, MacCallum RM, Roche DG, Ryan SJ, Schigel D, Vandegrift K, Watts M, Zaspel JM, Pawar S (2019) MIReAD, a minimum information standard for reporting arthropod abundance data. Scientific Data 6(1): 40. https://doi.org/10.1038/s41597-019-0042-5

Scholz AH, Freitag J, Lyal CHC, Sara R, Lucia Cepeda M, Cancio I, Sett S, Lee Hufton A, Abebaw Y, Bansal K, Benbouza H, Iddi Boga H, Brisse S, Bruford MW, Clissold H, Cochrane G, Coddington JA, Deletoille A-C, García-Cardona F, Hamer M, Hurtado-Ortiz R, Miano DW, Nicholson D, Oliveira G, Ospina Bravo C, Rohden F, Seberg O, Segelbacher G, Shouche Y, Sierra A, Karsch-Mizrachi I, da Silva J, Hautea DM, da Silva M, Suzuki M, Tesfaye K, Keambou Tiambo C, Tolley KA, Varshney R, Mercedes Zambrano M, Overmann J (2022) Multilateral benefit-sharing from digital sequence information will support both science and biodiversity conservation. Nature Communications 13(1): 1086. https://doi.org/10.1038/s41467-022-28594-0

Tedersoo L, Bahram M, Zinger L, Nilsson RH, Kennedy PG, Yang T, Anslan S, Mikryukov V (2022) Best practices in metabarcoding of fungi: From experimental design to results. Molecular Ecology 31(10): 2769–2795. https://doi.org/10.1111/mec.16460

Thalinger B, Deiner K, Harper LR, Rees HC, Blackman RC, Sint D, Traugott M, Goldberg CS, Bruce K (2021) A validation scale to determine the readiness of environmental DNA assays for routine species monitoring. Environmental DNA 3(4): 823–836. https://doi.org/10.1002/edn3.189

Wieczorek J, Bloom D, Guralnick R, Blum S, Döring M, Giovanni R, Robertson T, Vieglais D (2012) Darwin Core: An evolving community-developed biodiversity data standard. PLoS ONE 7(1): e29715. https://doi.org/10.1371/journal.pone.0029715

Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten J-W, da Silva Santos LB, Bourne PE, Bouwman J, Brookes AJ, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo CT, Finkers R, Gonzalez-Beltran A, Gray AJG, Groth P, Goble C, Grethe JS, Heringa J, 't Hoen PAC, Hooft R, Kuhn T, Kok R, Kok J, Lusher SJ, Martone ME, Mons A, Packer AL, Persson B, Rocca-Serra P, Roos M, van Schaik R, Sansone S-A, Schultes E, Sengstag T, Slater T, Strawn G, Swertz MA, Thompson M, van der Lei J, van Mulligen E, Velterop J, Waagmeester A, Wittenburg P, Wolstencroft K, Zhao J, Mons B (2016) The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data 3(1): 160018. https://doi.org/10.1038/sdata.2016.18

Yilmaz P, Kottmann R, Field D, Knight R, Cole JR, Amaral-Zettler L, Gilbert JA, Karsch-Mizrachi I, Johnston A, Cochrane G, Vaughan R, Hunter C, Park J, Morrison N, Rocca-Serra P, Sterk P, Arumugam M, Bailey M, Baumgartner L, Birren BW, Blaser MJ, Bonazzi V, Booth T, Bork P, Bushman FD, Buttigieg PL, Chain PSG, Charlson E, Costello EK, Huot-Creasy H, Dawyndt P, DeSantis T, Fierer N, Fuhrman JA, Gallery RE, Gevers D, Gibbs RA, Gil IS, Gonzalez A, Gordon JI, Guralnick R, Hankeln W, Highlander S, Hugenholtz P, Jansson J, Kau AL, Kelley ST, Kennedy J, Knights D, Koren O, Kuczynski J, Kyrpides N, Larsen R, Lauber CL, Legg T, Ley RE, Lozupone CA, Ludwig W, Lyons D, Maguire E, Methé BA, Meyer F, Muegge B, Nakielny S, Nelson KE, Nemergut D, Neufeld JD, Newbold LK, Oliver AE, Pace NR, Palanisamy G, Peplies J, Petrosino J, Proctor L, Pruesse E, Quast C, Raes J, Ratnasingham S, Ravel J, Relman DA, Assunta-Sansone S, Schloss PD, Schriml L, Sinha R, Smith MI, Sodergren E, Spor A, Stombaugh J, Tiedje JM, Ward DV, Weinstock GM, Wendel D, White O, Whiteley A, Wilke A, Wortman JR, Yatsunenko T, Glöckner FO (2011) Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIxS) specifications. Nature Biotechnology 29(5): 415–420. https://doi.org/10.1038/nbt.1823